

ABSTRACT

A video display device having a picture-in-picture (PIP) display, an audio input device, an image input device, and a processor. The device utilizes a combination of an audio indication and a related gesture from a user to control PIP display characteristics such as a position of the PIP within a display and the size of the PIP. A microphone captures the audio indication and the processor performs a recognition act to determine that a PIP control command is intended from the user. Thereafter, the camera captures an image or a series of images of the user including at least some portion of the user containing a gesture. The processor then identifies the gesture and affects a PIP display characteristic in response to the combined audio indication and gesture.